

This is an edited version of Mr. Kusunoki's three part article published in MJ magazine from Nov. 1996 through Dec. 1997. Translation by Yoshi and Irene Segoshi.

## Non-oversampling Digital filter-less DAC Concept by Ryohei Kusunoki

### To Confirm the Original 44.1kHz/16bit Format

It is exciting to create something new. Some people just grab a soldering iron, others deliberately start a simulation. Everyone seems to have his/her own approach. In my case, it starts with going back to the basics, research the history, and re-construct the whole picture in my mind. On starting this project I did research as many resources as I could put my hands on.

As the new generation of CD format is appearing on the horizon, I thought the basic concept should be "To Confirm the Original 44.1kHz/16bit Format". A CD in our hands has exactly the same data, every bit to bit, as to the one that left the studio. To recall this dreamy fact, the above theme would be quite appropriate. Any high-bit or high-sampling does not have its raison d'être unless it surpasses this level of accuracy.

### About Non-Oversampling

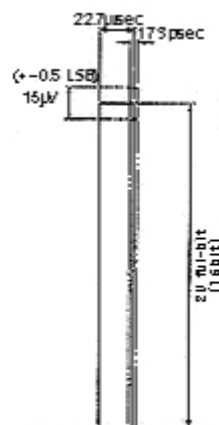
After examining the following two aspects, I came to a conclusion that 'it is quite difficult to carry out oversampling as theoretically under the current technology'.

#### 1) Oversampling and Jitter

There are two axes on digitizing the sound. The time axis and the amplitude axis. In case of CD, they are 44.1kHz and 16bit. In other words, we have to press in the amplitude data into one of the 16bit stage at every 22.7 s. That produces maximum of +0.5 LSB error, and the digital audio starts by accepting this error at the beginning. However, this error only concerns the amplitude axis and no amount of error was admitted on the time axis. Let me suppose that the accuracy of 16bit means how accurately the acoustic energy (time x amplitude) is transmitted by being distributed into each steps of 16bit. Then, by making the amplitude data more accurate, we can distribute the error onto the time axis.

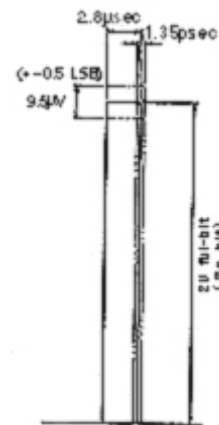
If we distribute 1/2 of the error,  $1 \div 44.1\text{kHz} \div 2^{16} \div 2 = 173 \text{ (ps)}$

This represents the maximum limit of the acceptable error (maximum limit of the jitter). (diagram 1)



[diagram1]

acceptable error of 44.1/16bit



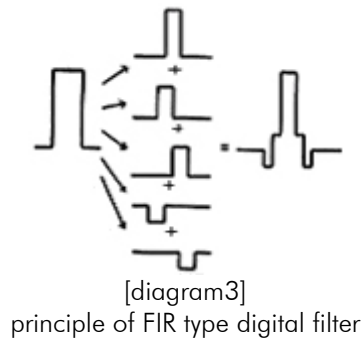
[diagram2]

acceptable error of 8 x sampling/20bit

All of the above is based on the basic sampling rate. When in 8 x oversampling and 20bit, that number would be 1.35ps (diagram 2). This is a totally impossible number to achieve for a separate type DAC which has to recover the clock by PLL. This means that under an average jitter environment, the oversampling can not operate theoretically, and lowers the accuracy within the operating field. In short, just by oversampling the original data, 16bit accuracy can not be satisfied anymore.

## 2) Oversampling and High-Bit

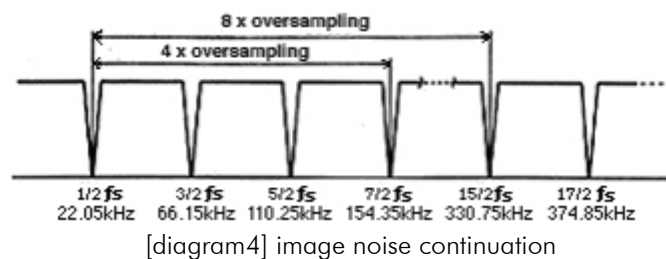
Originally, oversampling was developed to allow the use of an analog filter with gentler characteristics as a post-filter, and not to increase the amount of information. Many people still misunderstand this.



The principle of the most popular FIR type digital filter is to shift the original data and overlay them together, not to create an additional one (diagram 3). When it overlays the data by multiplying the coefficient to the original data, there appears new information below 16bit and to recover this finer information, we need a higher bit rate processing.

For example, in case of a high-performance digital filter SM5842, this processing is done in 32bit and the filter round them up to 20bit to the output, creating more errors in the re-quantizing process. Recently, this problem was dealt with and a filter was created which can produce 8 x sampling all at once. But even with that, as long as you can't output the internal word length as it is, there's no way you can prevent the errors to occur.

It may sound contrary, but if you take this error into account, 16bit without oversampling is more accurate than 8x-oversampling/20bit.



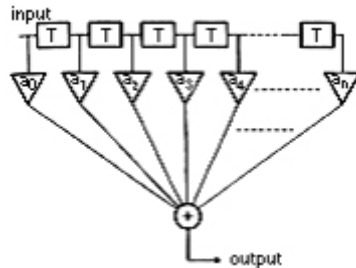
Then what is going to happen if you eliminate the oversampling process? Theoretically, the image noise will be repeated infinitely to higher frequencies (diagram 4), and a conventional answer would be 'it will sound awful'. Really? This has nothing to do with the "Shannon's theorem", nor do I intend to challenge that. Shannon's theorem considers a sampling theory on transmitting an information. I am talking about the perception of the information. That is, if I must say, "the limitation of our auditory sense is a powerful low-pass-filter and the Shannon's theorem is satisfied at the echelon of human auditory perception." My challenge is rather toward those who listen to the sound through theories and oscilloscopes.

Another way of thinking is that, even if humans can't hear it, the equipment that follows can and will be affected by it.

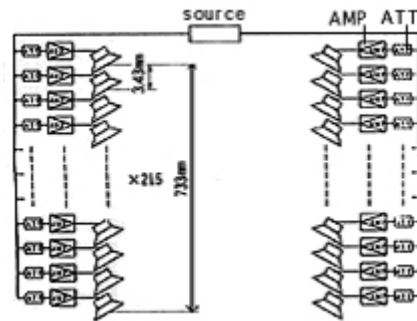
However, 8x-oversampling/digital-filter can only cut off the frequencies between 22.05kHz and 330kHz. Everything beyond 330kHz is all coming through untouched, meaning the degree of effect is determined by how the said equipment reacts to the ingredients beyond 330kHz. My guess is, if 100kHz sign wave comes through, there won't be any problem.

## Problems of the Digital Filter

The diagram 5 shows the principle of the most popular FIR type digital-filter. The "T" represents a delay circuit for each sampling interval, "a" is for the coefficient multiplier, and "+" is an adder. After delaying the input data, it multiplies with the coefficient, and this process is repeated n times. This 'n' is called the number of taps. The more taps it has, the higher the performance of the filter is supposed to be. The delay mentioned above is not that of a calculating time, but more like a waiting time until the next data arrives.



T: a delay circuit  
for each sampling interval  
a: coefficient multiplier  
+: adder  
[diagram5]  
FIR type digital filter



[diagram6]  
FIR type digital filter (in case of SM5842)

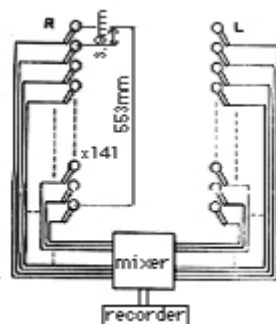
It is rather hard to understand this diagram instinctively. It didn't hit home with me, either. But, one day, it occur to me to replace it with the equivalent of the reproducing hardware system. (diagram 6). The delay circuit is replaced with that of the delay of speed of sound, the multipliers with the attenuators, and the adding is synthesized in the space. The number of the speakers corresponds to that of taps. The diagram shows, as an example, the computation of CD data through the high-performance digital-filter SM5842. The accompanied numbers are the actual sizes in the space when replaced with the hard-ware. Since the sampling frequency of CD is 44.1kHz, each delay time for the 1 x sampling is 22.ms per tap. To achieve 8 x sampling, SM5842 repeats 2 x sampling three times, and each step incorporates the taps of, 169 degrees for 2 x, 29 degrees for 4 x, and 17 degrees for 8 x. The accumulated delay of each step becomes, 1.92ms, 0.16ms, and 0.05ms: total of 2.13ms.

Our auditory sense does the frequency analysis at every 2ms interval, and 2.13ms of delay can be caught by our ear.

If the speed of sound is 346m/s, the total length of the row of speakers becomes 737mm. ( In the diagram, the distance between each speaker is presented by the total delay divided by the total number of taps.)

Now, you can imagine what kind of sound will result from such a system. All the notes coming from the speakers before and behind, will mix, intervene with each other, and spread. I would like to express this expansion of the sound over the time axis as a "diffusion of sound coherence". For example, if an attack of a piano note was not clear enough, as if the felt on the hammer became thicker, you might be hearing this "diffusion of sound coherence".

It may sound contrary, but if you take this error into account, 16bit without oversampling is more accurate than 8x-oversampling/20bit.



[diagram7]  
in case SM5815A  
is used in 1/2 decimation

The diagram 7 indicates the diaphragm 5 replaced with a recording hardware. If you ever felt the digital recording somewhat lacking a core of the sound, please examine this illustration carefully. In a way, one point recording using digital filter is so much nonsense. The time will come in the near future when the performance of a digital filter will be evaluated not only by its cut-off characteristics but also how small a number of taps it has. If the digital filter is a necessary evil, we have to make sure to limit the total delay within 2ms throughout the recording and playback so that it won't be caught by human auditory sense.

## The Sound of Non-oversampling

We can control the "diffusion of sound coherence" only by constructing it with smaller number of taps. From that aspect, Wadia's decoding computer (13 taps) or Luxman's former fluency DAC, DA-07 (3 taps) are considered to be excellent machines. They both received (Wadia still does) outstanding appraisals at the time for their sensual representation of the sound. The sound of non-oversampling DAC is on the extension of these machines, and theoretically, it can exceed those achievements.

The difference between the non-oversampling DAC and the conventional DAC with the digital filter lies whether you attach importance on the accuracy in the time domain or in the frequency domain. In other words, whether you choose the musical performance or the quality of a sound. This trade-off line defines the boundary of the current digital audio format.

A natural, stress-free sound that communicates the musicians' intention directly to you. That is the sound of non-oversampling DAC. The feel of this sound is closer to that of analog reproduction.

## Introducing Non-PLL clock

We can still hear the characteristics of each different transport even after lowering the jitter sensitivity to a minimum by non-oversampling. This is an incomplete section of today's digital audio format. The fundamental advantage of being digital, that its quality does not depend on the conveying form, falls completely short here. That is because we have to create the time axis from the incoming data by PLL at the receiving end. This is a flaw inevitable to the current DAI format which requires word-synchronization. It is often misunderstood as if the time axis is digital, instead of analog, because it receives discrete value after sampling, but actually, it is completely analog. When the time axis is distorted, the analog wave form is distorted with it.

Then what is going to happen if we read it by its own clock at the receiving side? Unfortunately, the word-synch can not be kept and the waveform will be broken into shreds. I tried to re-clock it with a separate non-PLL-clock after once it's locked with PLL and reproduced (diagram 8). By doing this, the fluctuations within 1 clock of the PLL are completely absorbed and not transferred to the reproduced waveform. But any fluctuations beyond 1 clock, even if it's only 1ps, are magnified into that of 1 clock. This happens often, because of the frequency difference between the PLL clock and non-PLL clock. In case of this 50MHz non-PLL clock, the amount is 20ns per every 0.1ms. This is, after all, more than 100 times of the 16bit criterion.



1. Non-PLL clock 50MHz
2. PLL clock 2.8224MHz(44.1kHz x 64fs)
3. re-clock pulse  
[diagram8]  
Non-PLL re-clock

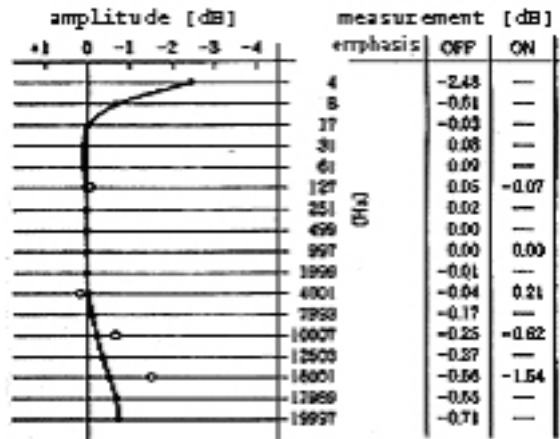
How does it sound? Are the notes broken or jumping around and unbearable to listen to? Somehow it does not, and not only that; it generates an extremely realistic sound field. A certain acoustic atmosphere envelops the room making you feel like you are on the same floor with the performers, communicating the tensions and relaxation among musicians to you.

This experience made me wonder if human ears are somewhat insensible to jitter. Whether it has a large or a small amount of jitter is not really an issue. The real problem is the constant fluctuation of the time axis caused by PLL. What is more important is the structure, including the time axis, of jitter, rather than the amount of it.

However, we can still hear the different characteristics of transports. I suspect that we may be hearing an effect of the original jitter detected by the non-PLL clock and entwined around the beat component.

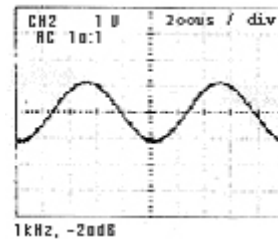
## Measurements

The diagram 9 indicates the frequency characteristics (with the emphasis on and off). It looks a tube amplifier measurements. The roll off at lower frequencies seems caused by narrow frequency range of the analyzer itself, because I get the same response with whatever CD player I'm measuring (the manufacturer's spec sheet claims 20Hz~100kHz). The fall at the higher frequencies are caused by the aperture effect.



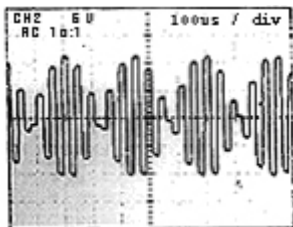
DENNON Professional Test CD CCCC- 75084 - 86  
 ● emphasis OFF DISC-1 Track 20 - 24  
 ○ emphasis ON DISC-1 Track 26 - 30

[diagram9]  
Frequency

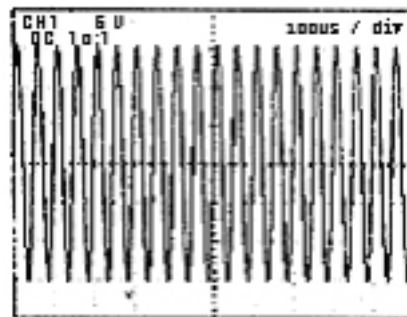


[diagram10]  
Non-oversampling DAC

The diagram 10 indicates 1kHz sign wave, -20db. For a comparison, the same waveform through a conventional DAC with digital-filter is shown on the diagram 11. The obvious notches that remained are all formed by components beyond 20kHz and could not be detected by human ear.

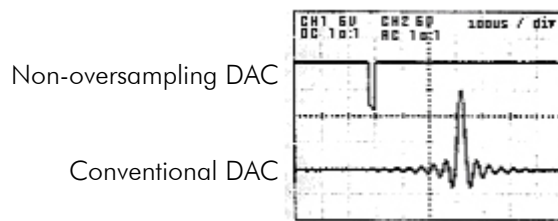


[diagram12]  
Non-oversampling DAC



[diagram13]  
Conventional DAC

20kHz, 0db is indicated in diagram 12, and that of the comparing DAC in diagram 13. In diagram12, it seems like 22kHz square wave is under amplitude modulation at 4kHz, and no 20kHz can be seen. I am not sure if this is perceived as 20kHz when it is filtered through our auditory sense. I would like to hear opinions from psycho-acoustic professionals. Considering the limit of impairment perception of humans (around 200Hz), however, this amplitude modulation of 4kHz does not need to be worried about at all.



[diagram14]  
Full-bit impulse

The impulse response is shown at the diagram 14. The comparing DAC's response is juxtaposed in the same diagram. The one with the impulse shown downward is that of the non-oversampling DAC. With passive I/V conversion, unless you invert the data somewhere along the way, it comes out as opposite phase. While the comparing DAC shows a familiar waveform, non-oversampling DAC indicates an excellent pulse response. The slant of the top (bottom?) is caused by a low pass filter (160kHz at the time of the measurement). The pre-post echo shown at the bottom picture indicates the "diffusion of sound coherence". I am not saying that you hear this echo as it is, but suggesting that the process which produce this waveform has a problem in itself. If you examine this waveform more closely, undulations of longer term should be observed ahead and behind each echo.

The diagrams here are of Mr. Kusunoki's proto-type DAC and not of the PROGRESION converter.

#### A Comment on the New Formats

The relationship between the sound and the measurement still remain mysterious. You cannot achieve good sound just by competing on the numbers of zero over the distortion factor nor by excessively extending the frequency range. However, in the next generation digital format offered today, the selling points for better sound are quantizing bit numbers and sampling frequency rates. It only means lowering of distortions and extension of frequency range.

The appearance of CD was an epoch-making event as a new format to follow LP. It delivered the sound of the master tape to our listening room. It was a crystallization of efforts of the engineers of that time.

Compared to that, the new generation CDs offered today only concern raising the data rate; something similar to the idea of EL cassettes. The life span of the format will be a very short one. What we need is to accurately understand the merits and demerits of the current format and create the new one that pertinently matches our auditory sense.